# HW1

Use the data set in case study 2, and utilize R to answer the following questions.

Note: you need to take snapshots of your R coding & outputs, also use your texts to answer the following questions (if necessary). Just like the example given in our slides. Upload a single PDF document as your submission

1. Load dataset into R and show the column names, example of values, and the size of the data [10]
2. Return a list of records with columns <day, hours, mail, cert>, where day is Friday, and the number of mails is larger than 7000 [10]
3. Change the column name from “cert” to “certificate” [5]
4. Use descriptive statistics to understand the column “day”. More specifically, return class frequency, and class relative frequency. Also visualize this column by using bar graph (use class relative frequency as y-axis) and pie chart [20]
5. Use descriptive statistics to understand the column “certificate”. More specifically, we want to get q1, q2, q3, average value, variance value. Also, visualize this variable by using histogram, and interpret your histogram [20]
6. Use descriptive statistics to understand the column “misc”. Visualize it by using boxplot, and interpret your boxplot [20]
7. Visualize the column “tickets” by using probability curve, interpret it [15]